

# Cooperative Content Dissemination and Offloading in Heterogeneous Mobile Networks

Guoqiang Mao, *Senior Member, IEEE*, Zijie Zhang, *Member, IEEE*, and Brian D. O. Anderson, *Life Fellow, IEEE*

**Abstract**—To cope with an exponentially increasing demand on mobile data traffic in cellular networks, high-capacity WiFi and device-to-device networks can be exploited as a complementary means to offload and reduce the traffic load of cellular networks. This paper proposes a novel cooperative content dissemination strategy for a heterogeneous network consisting of different types of devices, such as WiFi access points, smart phones and intelligent vehicles. The proposed strategy offloads a significant proportion of data traffic from cellular networks to WiFi or device-to-device networks using ad hoc connections that emerge opportunistically when mobile devices move and meet one another. The strategy is particularly suited for dissemination of delay tolerant content. Detailed analysis is provided for the content dissemination process in heterogeneous mobile networks adopting the proposed strategy. On that basis, the optimal parameter settings for the content dissemination strategy are discussed. Simulation and numerical results show that the proposed strategy significantly reduces the amount of data traffic for cellular networks while guaranteeing the successful delivery of content.

**Index Terms**—content dissemination; data offloading; mobile network; heterogeneous network; epidemic broadcast; erasure coding

## I. INTRODUCTION

With the development of wireless communication technologies and the increase in the number of mobile devices such as smart phones, tablet PCs and intelligent vehicles, there is an explosive demand on the data traffic in mobile cellular networks [1]–[3]. This paper investigates a content dissemination strategy for heterogeneous mobile networks where a portion of data traffic is offloaded from cellular networks to high-capacity and low-cost complementary networks, including WiFi, vehicular and device-to-device networks, to ease the burden of cellular networks.

Content dissemination in mobile networks has become an increasingly important and challenging task. It is predicted that an average mobile broadband user will soon consume 7 GB

G. Mao is with the School of Computing and Communications, The University of Technology Sydney. He is also with National ICT Australia, Beijing University of Posts and Telecommunications and Huazhong University of Science and Technology. E-mail: g.mao@ieee.org.

Z. Zhang's work was mainly undertaken while he was associated with the School of Electrical and Information Engineering, University of Sydney and National ICT Australia. E-mail: zijie@ieee.org.

B. D. O. Anderson is with the Research School of Engineering, Australian National University and National ICT Australia. E-mail: brian.anderson@anu.edu.au.

This research is supported by Australian Research Council (ARC) Discovery projects DP110100538 and DP120102030 and Chinese National Science Foundation project 61428102.

Copyright (c) 2015 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

of traffic per month, which is 5.4 times more than the current value [4]. Usually in cellular networks, a user downloads content directly from a Content Provider's (CP) server over a Communication Service Provider's (CSP) network. Due to limited bandwidth of cellular networks, an increasing number of base stations are deployed by CSPs to increase the capacity of cellular networks to fulfil the aforementioned explosive data traffic demand. In addition to the capacity issue, downloading content through CSP networks incurs a cost, which must be borne either by the end users or by the CPs [5].

To solve the challenges in both capacity and cost, there is intensive research on offloading mobile data traffic from cellular networks, which rely on congested and expensive licensed frequency bands, to complementary networks, such as WiFi and device-to-device networks, which typically use free unlicensed frequency spectrum [4], [6].

In addition to complementary networks, the mobility of users can also be exploited to assist the content dissemination. When users carry mobile devices physically while walking around university campus, shopping centres or travelling by taxis or buses, the content in their mobile devices also move with them without consuming any bandwidth. Together with device-to-device communication technologies, mobility of users offers an alternative way to transport delay-tolerant content efficiently [1].

A large proportion of the content delivered by CSPs over mobile networks is delay-tolerant content, like videos, newspapers and weather reports [7]. Consider a typical example of delay-tolerant content dissemination of videos or newspapers via pedestrians. The mobile device carried by a pedestrian may receive the content from a cellular base station or a WiFi access point. As the pedestrian moves, the content moves with the pedestrian. When the pedestrian encounters another pedestrian(s) interested in the content, the pedestrian may pass the content through device-to-device connection between them. In this way, the content originally carried via cellular networks is offloaded to the WiFi and device-to-device networks. Consider another example of delay-tolerant content dissemination in taxis or buses. The dissemination of multimedia content (e.g. video clips) to the vehicles can be costly using cellular networks. In contrast, vehicular networks are an important class of mobile networks, which have attracted increasing interests recently. Vehicles are highly mobile and can communicate with each other or with nearby roadside units using the dedicated short range communication (DSRC) radio [8]. These features make vehicular networks a suitable candidate for data offloading too, i.e. using a complementary network and device mobility to ease the burden of cellular

networks. In both examples, the content is offloaded from cellular networks to other complementary networks. This reduces the traffic load of cellular networks (and boosts the capacity of cellular networks in the sense that more users can be served) and also provides a higher data rate for the users. The expense is that content offloading will incur higher delay compared with direct transmission using cellular networks.

This paper investigates content dissemination and offloading in heterogeneous mobile networks, which reduces the traffic load of cellular networks while guaranteeing the successful delivery of content.

More specifically, the following contributions are made in the paper:

- 1) a cooperative content dissemination strategy is proposed, which exploits all three data dissemination methods, i.e. cellular networks, complementary networks and device mobility, to disseminate content;
- 2) an analytical framework is proposed for studying the information dissemination process, taking into account all three data dissemination methods. Analytical results are presented to stochastically characterise the content dissemination process, considering the heterogeneity in the devices, in terms of mobility and transmission capability. On that basis, the reduction in the amount of data traffic in cellular networks is calculated;
- 3) based on the above results, the optimal parameter settings of the content dissemination strategy are discussed. Simulation and numerical results show that the proposed strategy can effectively reduce the traffic load of the cellular network while guaranteeing the successful delivery of all content.

The rest of this paper is organised as follows: Section II reviews related work. Section III introduces the system model, including the content dissemination strategy. Section IV presents the analysis of the content dissemination process, whose optimal design is studied in Section V. Section VI validates the analysis using simulations. Finally Section VII concludes this paper and proposes possible future work.

## II. RELATED WORK

The explosive growth of mobile data traffic has attracted significant attention from both academia and industry. A number of innovative solutions have emerged to offload data traffic and reduce the load of cellular networks, as well as increase the capacity of cellular networks [3]. Some of the key technologies [7] include small cell technology and using WiFi networks and device-to-device networks for mobile data offloading. This paper focuses on the latter technology.

Recent research [4] has shown that WiFi networks have already carried and offloaded a large amount of mobile data traffic. When a device enters into a WiFi-covered area, it can switch its data traffic from a cellular network to a WiFi network [9] to reduce cost and traffic load of the cellular network. A major issue in WiFi offloading is the optimum deployment of WiFi access points (APs) for efficient offloading. Bulut et al. [9] compared different methods of deploying WiFi APs for efficient offloading of mobile data traffic. They

also proposed a greedy approach that can achieve a high offloading efficiency, i.e. a higher percentage of data traffic is offloaded to WiFi networks, compared with existing methods like Sequential or Hot-Zone. Moreover, instead of relying solely on existing WiFi APs, recent development of technology has also allowed mobile devices to become a virtual WiFi access points (a.k.a. WiFi Tethering [10]), so that devices can communicate with one another opportunistically without relying on infrastructure. In this way, devices can *cooperate* with one another to disseminate content.

Using a typical cooperative content dissemination strategy [7]; the service providers first deliver the content to only a small set of users, then these users can further disseminate the content to other subscribed users when their mobile devices are in proximity and can communicate using WiFi tethering or Bluetooth technology. It is obvious that entirely relying on the opportunistic content dissemination cannot a priori guarantee the delivery of content. This paper proposes a mechanism that can provide guarantee on the delivery of content by combining the use of cellular networks and opportunistic networks.

A further point of differentiation of this paper is as follows. Existing work on cooperative mobile data offloading (e.g. [1], [11]) mainly focuses on the design and selection of helpers, viz. some special mobile devices selected to help the content provider to deliver messages to other mobile devices using ad hoc connections. A recent work for example of Li et al. [1] considered the impact of the buffer size constraint on the amount of data offloaded by helpers. Specifically, assuming Poissonly distributed inter-contact time interval between mobile users, Li et al. [1] proposed three algorithms for storage allocation of helpers to maximise the system offloading utility function taking into account the constraints on the helpers' buffer-size. In contrast, this paper considers the case where every node can be a potential helper and issues such as buffer size become of much less importance in the scenarios considered in the paper.

As vehicular networks are an important category of mobile networks, content dissemination in vehicular networks has also been investigated [2], [6]. Particularly, wireless access through vehicle-to-roadside communications can be utilised in public transportation vehicles for streaming applications, e.g., video and interactive advertisement. As pointed out in [12], it is challenging to develop an efficient wireless access scheme that minimises the cost of wireless connectivity for downloading data. There are also a number of empirical studies on content dissemination methods for vehicular networks based on simulations only [13]. In this paper, we propose a cooperative content dissemination strategy backed by analytical studies and the strategy is particularly suitable for vehicular networks.

Research in the area of content caching is also relevant to this work. Taghizadeh et al. proposed a cooperative caching strategy in [5]. Specifically, they introduced a novel peer-to-peer rebate mechanism to encourage the users to cache previously downloaded content and to share it with other users, taking into account the selfish behaviour of users. This paper not only takes into account the willingness of user cooperation but also considers the impact of heterogeneity in device mobility on the content dissemination process.

### III. SYSTEM MODEL

#### A. Network model

We consider a heterogeneous network [14]–[16] consisting of  $N$  nodes. Without loss of generality, suppose that there are  $H \in \{1, 2, \dots, N\}$  types of nodes, e.g. static WiFi APs, mobile devices carried by pedestrians or vehicles, nodes with different social characteristics [15]. Let  $N_h$  be the total number of nodes of type  $h$ :  $\sum_{h=1}^H N_h = N$ .

For our work, it is important to consider spatial distributions of nodes. We shall adopt an assumption that is obviously never exactly fulfilled in practice, but is nevertheless indicative of what will happen in a real network. Suppose that at some initial time instant  $t = 0$ , the  $N$  nodes are independently, randomly and uniformly distributed on a torus  $(0, L]^2$  [17]. It follows that the nodes' density is  $\lambda = N/L^2$ . The consideration of a torus allows us to ignore the boundary effect and focuses on the impact of main parameters that affect the design and performance of the content dissemination strategy. After time  $t = 0$ , these nodes start to move independently according to some mobility models [18]. As commonly done in the field, we assume that the nodes' mobility is such that the spatial distribution of nodes is stationary and ergodic with stationary uniform distribution on the torus. Our theoretical analysis can be applied to a class of mobility models [19] under which the spatial distribution of nodes is stationary and ergodic, and does not rely on the use of a particular mobility model. As shown in [19], a number of mobility models have this property. On the other hand, we acknowledge that under some mobility models, e.g. the well known random waypoint model [20], the spatial distribution of nodes is no longer stationary and ergodic. Particularly under the random waypoint model, even if the nodes are initially uniformly distributed on a square, as time goes on, the spatial distribution of nodes will change such that nodes will be concentrated around the centre of the square [20]. Our analysis cannot be applied to these mobility models. In Section VI, we use simulations to study the performance of the proposed content dissemination strategy under the random waypoint model.

Note that in many scenarios encountered in real applications, the spatial distribution of nodes is clustered instead of being uniform. We limit our theoretical analysis to uniformly distributed nodes for analytical tractability. Particularly, almost all existing related studies are based on uniform or Poisson node distribution. The main reason is that the theoretical tools used in the studies are mainly based on stochastic geometry or continuum percolation theory where results are only available for uniform or Poisson node distribution. Realising the limitation of uniform or Poisson distribution, there have been attempts at establishing mathematical tools for more general node distribution [21]–[24]. So far, these attempts have had only limited success and we are yet to develop the theoretical tools for analysing more general node distribution. The simulation presented in Section VI assuming the random waypoint model sheds insight on the performance of the proposed dissemination strategy under clustered node distributions because under the random waypoint mobility model, even when the initial distribution of nodes is uniform, as time

evolves, the spatial distribution of nodes becomes clustered [20].

In some parts of the paper, we do need to specialise to consider a particular mobility model. For example, when using simulations to evaluate the performance of the proposed content dissemination strategy, a particular mobility model has to be adopted. In this case, we consider a widely used mobility model called the random direction model (RDM) [18]. Specifically, under the model, at time  $t = 0$ , every node chooses its direction independently and uniformly in  $[0, 2\pi)$ , and then moves thereafter at a constant speed  $V$ . The speed may be different for nodes of different types. Obviously  $V = 0$  for stationary nodes. It has been shown by Nain et al. in [19] that under the aforementioned model, at any time instant  $t \geq 0$ , the spatial distribution of nodes is stationary and follows the same uniform distribution.

#### B. Wireless communication model

As introduced earlier, we consider two interacting types of wireless networks: the cellular network and the complementary network; the latter can comprise mobile and stationary nodes.

It is assumed that within any sufficiently long time interval, every node is able to be directly connected to at least one cellular base station (BS) at a time instant during the interval and therefore is able to receive packets from the cellular network within the time interval. If a mobile node temporarily moves outside the coverage of the cellular network, it is able to be directly connected to the cellular network when it moves back into the coverage area. As introduced earlier, the bandwidth of BSs is limited and a large number of packet transmissions via cellular networks can be monetarily expensive and inefficient for the mobile users and the content providers.

In the complementary network, devices communicate with one another via ad hoc connections using device-to-device communication technologies such as Bluetooth, WiFi or DSRC [8], [10]. These ad hoc connections are usually of high capacity, and can be exploited by mobile devices to cooperatively share content of common interest. The challenge is that ad hoc connections usually have a relatively short range (e.g. 10 meters using Bluetooth, 20 meters using Wi-Fi Tethering [10] or 250 meters using DSRC [8]).

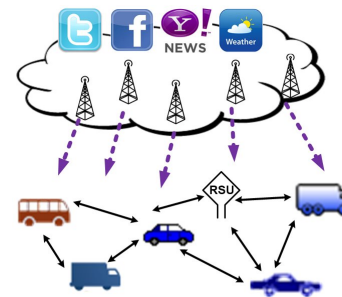


Fig. 1. Illustration of the system model for cooperative content dissemination in mobile networks.

Due to the limited communication range, the ad hoc connection between two nodes only appears when they move

close to each other. Considering a commonly used model, called the *unit disk connection model* (UDM), two nodes are directly connected iff the Euclidean distance between them is not larger than the radio range  $r_0$ . Adopting this commonly used connection model, we say that two nodes *meet* each other when they move into the radio range of each other.

Define the *inter-meeting time* of two nodes as the time interval between two consecutive meetings of two nodes. There have been a number of studies on the probability distribution of the inter-meeting time. In particular, Cai and Eun [25] analytically proved that for two nodes moving in a finite area (with reflective or wrapping boundary) under random waypoint or random walk mobility models<sup>1</sup>, their inter-meeting time has an exponential distribution, whereas the inter-meeting time of nodes moving in an infinite area follows a power-law distribution. This paper assume that the inter-meeting time follows an exponential distribution. In the Appendix, we show that for nodes moving according to the random direction model, their inter-meeting time follows an exponential distribution. Furthermore, the analysis presented in this paper is directly applicable to other inter-meeting time distributions as shown in Section IV-A.

### C. Cooperative content dissemination strategy

This section describes a cooperative content dissemination strategy which makes use of both the cellular network and the complementary network.

1) *Cooperative content dissemination strategy*: Consider that a content provider has  $M$  messages (e.g. advertisements, news updates or a video clip) to deliver to  $N$  nodes. The content dissemination process has three phases: initial phase, sharing phase and complement phase.

Phase 1 Initially at time  $t = 0$ , the cellular BSs transmit (a.k.a. push)  $\beta$  packets to  $\beta$  different nodes via the cellular network. The content in these packets depends on the coding scheme to be described in Section III-C2. On the one hand, the relationship between  $\beta$  and  $M$ , particularly the maximum value of  $\beta$ , depends on the coding scheme, which will be discussed in Section III-C2. On the other hand, as later shown in Section V, the value of  $\beta$  can also be optimally chosen to minimise the traffic load of cellular networks.

Phase 2 Then the network enters into the *sharing phase*, where the nodes broadcast their received packets using the Susceptible-Infected-Recovered (SIR) epidemic scheme, to be described in Section III-C3. Some of the nodes may be APs or stationary.

Phase 3 At time  $\mathcal{T}_{end}$  (called deadline), the sharing phase stops and the network enters into the *complement phase*, where every node requests (a.k.a. pulls) the remaining packets required to decode all  $M$  messages from BSs when it is within the coverage of the cellular network, where the *remaining packets* required to decode the messages are determined by the coding scheme, to be described in Section III-C2. The value

of  $\mathcal{T}_{end}$  has a significant impact on the performance of the content dissemination strategy and will be discussed later.

The main objective in the design of the content dissemination strategy is to minimise the total number of packets requiring to be transmitted through cellular networks, which include the packets transmitted in the initial phase and in the complement phase, while guaranteeing the successful delivery of the content.

2) *Coding scheme*: Erasure codes have been shown to have a good tolerance to delay and packet losses when reconstructing original messages from packets transmitted through unreliable networks [26], [27].

This paper considers a simple erasure coding scheme. Specifically, the  $M$  messages are encoded into  $\beta$  coded packets. Furthermore, we assume that the Galois field [26], [27], denoted by  $\text{GF}(2^\phi)$ , used in the encoding process is large enough so that the content provider can generate  $\beta$  linear-independently coded packets. Particularly, the Galois field  $\text{GF}(2^\phi)$  needs to be chosen such that  $2^\phi \geq \beta$ . It follows that each node needs to acquire  $M$  distinct coded packets to reconstruct all the  $M$  messages.

Consequently, if a node receives  $B < M$  coded packets by the end of the sharing phase, then it needs to pull  $M - B$  coded packets in the complement phase via cellular networks in order to decode all  $M$  messages.

For benchmarking, we also consider the case that no coding technique is employed. In this case, the BSs transmit  $\hat{\beta}_m \in \{0, 1, \dots, \beta\}$  copies of the  $m^{\text{th}}$  message in the first phase for  $m = \{1, 2, \dots, M\}$  and  $\sum_{m=1}^M \hat{\beta}_m = \beta$ . Then each node needs to receive at least one copy of each message.

3) *Epidemic sharing scheme*: In the sharing phase, the  $\beta$  packets are shared among  $N$  nodes via ad hoc connections using a Susceptible-Infected-Recovered (SIR) epidemic sharing scheme.

Without loss of generality, consider the epidemic sharing of an arbitrary packet, say packet  $j$ . Using a classic SIR scheme, a node can be in any of the following three states regarding the packet  $j$ : the node that has never received the packet  $j$  is in the state of susceptible ( $S_j$ ). A susceptible node goes into the state of infected and infectious ( $I_j$ ) immediately after receiving the packet  $j$ . The node in state  $I_j$  keeps transmitting the packet  $j$  to every node it meets for a certain time period, which is referred to as the *active period*. Denote by  $\tau_h$  the length of the active period of a type- $h$  node. Parameter  $\tau_h$  is a pre-determined value, which is the same for all nodes of the same type. After the active period, the node recovers and enters into state  $R_j$ . A recovered node stops transmitting and receiving the packet.

The nodes that have received the packet  $j$  are referred to as the *informed nodes of packet  $j$* .

Note that the length of the active period  $\tau_h$  is determined by the capability and level of cooperation of each type of nodes. For example, a pre-installed WiFi AP can have a significantly larger value of  $\tau_h$  compared with other mobile devices that are powered by battery. Furthermore, the value of  $\tau_h$  for a mobile device can be tuned by introducing some incentives (e.g. a lower subscription fee or some rewards) to the mobile users

<sup>1</sup>Note that the random direction model introduced in Section III-A is a special case of the random walk model.

[11], so that the mobile users are willing to share more packets with other users. It is well known that nodes can be selfish and therefore incentives for cooperative communications is an important topic that has been extensively studied in the context of ad hoc networks and cooperative communications. It is beyond the scope of this paper to study the impact of incentive and selfish behaviour of nodes on cooperative communications. We refer interested readers to [11], [28] and references therein for further discussion on the topic.

The popularity of the content and the user preference may also affect the willingness of nodes to participate in the cooperative communications and to forward the received packet. A simple model to capture this effect is to add a parameter  $p$  such that instead of forwarding the received packet indiscriminately, the node receiving the packet may forward the packet with probability  $p$  where the value of  $p$  is adjusted according to the popularity of the content and the user preference. Our theoretical analysis can be readily extended to incorporate such a parameter into the packet propagation process. This will effectively reduce the density of cooperating nodes in the cooperative communications such that the total number of nodes participating in the cooperative content dissemination reduces from  $N$  to  $pN$ .

Note that  $\mathcal{T}_{end}$  is an important parameter for the epidemic sharing process and the optimisation of  $\mathcal{T}_{end}$  is an important and interesting problem to study. When the length of the sharing phase  $\mathcal{T}_{end}$  is sufficiently large, the epidemic sharing process stops naturally (i.e. reaches the *steady state*) when, for all the packets, there is no infectious node. We are particularly interested in the case when  $\mathcal{T}_{end}$  is large because it is suitable for delay-tolerant content disseminations and minimises the data traffic load of the cellular network by fully utilising the complementary networks to share content. On the other hand, choosing a smaller value of  $\mathcal{T}_{end}$  may help to reduce the delay incurred during the data dissemination process. When  $\mathcal{T}_{end}$  is small however, the system will be in a transient state and the associated transient state analysis becomes notoriously difficult. We leave it for our future work to consider the optimisation of  $\mathcal{T}_{end}$ .

Furthermore this paper considers a large network where  $N \gg M$ . In the next section, when we consider an *asymptotic* network with  $N \rightarrow \infty$ , we increase the network area  $L \rightarrow \infty$  while keeping the density of every type of nodes unchanged. That is, we adopt the widely used extended network model in our analysis.

#### IV. ANALYSIS OF THE CONTENT DISSEMINATION PROCESS

The main challenge in the analysis of the content dissemination process is the analysis of the packet propagation process in the sharing phase. This section first analyses the propagation process of single packet and then generalises to the entire set of packets.

##### A. Characterising the ad hoc connections

Recall that there are  $H$  types of nodes in the network and nodes move independently of one another. This sub-section

considers the inter-meeting process of a randomly chosen type- $h$  node and a randomly chosen type- $k$  node.

The inter-meeting time of two nodes is the time interval between consecutive meetings of two nodes. Denote by  $T_{h,k}$  the inter-meeting time between a randomly-chosen type- $h$  node and a randomly-chosen type- $k$  node, for  $h, k \in \{1, 2, \dots, H\}$ .

When  $T_{h,k}$  follows an exponential distribution with mean  $\lambda_{h,k}$ , the pdf of the inter-meeting time can be written as

$$\Pr(T_{h,k} = t) = \lambda_{h,k} \exp(-\lambda_{h,k}t). \quad (1)$$

In the Appendix, it is shown that the distribution of the inter-meeting time for nodes moving following the random direction model is an exponential distribution. Furthermore, the exponential inter-meeting time distribution has been widely used in a large number of research (e.g. [29]–[31]).

It follows from Eq. 1 that the probability that the type- $h$  node meets the type- $k$  node during the active period of the type- $h$  node is

$$\gamma_{h,k} = \int_0^{\tau_h} \Pr(T_{h,k} = t) dt = 1 - \exp(-\lambda_{h,k}\tau_h). \quad (2)$$

If the inter-meeting time follows a distribution other than the exponential distribution, a similar method can be used to calculate  $\gamma_{h,k}$ . Thus our analysis does not rely on the inter-meeting time following an exponential distribution.

Define a symmetric matrix  $\Xi \triangleq \{\gamma_{h,k}; h, k = 1, 2, \dots, H\}$ . The matrix stochastically characterises the ad hoc connections between every pair of devices in the network. It is interesting to note that in the special case when  $H = N$ , the matrix  $\Xi$  is equivalent to the (probabilistic) adjacency matrix of the network. Furthermore, the case where  $H = 1$  corresponds to a network where nodes have an identical mobility and the same active period. When  $H = 2$ ,  $V_1 = 0$ ,  $V_2 \neq 0$ ,  $\gamma_{1,1} = 0$ ,  $\gamma_{2,2} = 0$ ,  $\gamma_{1,2} > 0$  and  $\gamma_{2,1} > 0$ , it corresponds to the traditional non-cooperative data offloading networks where mobile devices (type-2 nodes) can only receive data packets from cellular BSs or WiFi APs (type-1 nodes). Besides the above examples, more interesting cases, including a network with more than two types of nodes, are studied later in Section VI.

Note that  $\gamma_{h,k}$  is determined by the nodes' mobility model and the length of the active period. There are two assumptions imposed on  $\gamma_{h,k}$  so that the analysis becomes non-trivial:

**Assumption 1** (Strictly positive). Let  $\gamma_{h,k}^n$  be the  $(h, k)^{th}$  element of  $\Xi^n$  for  $n \in \{1, 2, 3, \dots\}$ . There exists a finite value of  $n$  such that  $\gamma_{h,k}^n > 0$  for all  $h, k = 1, 2, \dots, H$ . In this case, the matrix  $\Xi$  is called *strictly positive* [32, Chapter V].

The above assumption means that the underlying graph of the network over time  $[0, \infty)$  is connected, where the *underlying graph* of a network over a time interval  $[0, t)$  is defined [29] to be an undirected graph where each vertex uniquely represents a node and each edge uniquely represents a link between two nodes, if the link exists between the two associate nodes at *any* time instant during  $[0, t)$ .

**Assumption 2** (Finite number of connections). When considering a large network, i.e. increase the area size  $L^2$  while keeping nodes' density unchanged, the communication ranges

of all nodes are decreased so that  $N_k\gamma_{h,k}$  is a finite constant for all  $h, k = 1, 2, \dots, H$ .

The above assumption lets us avoid the trivial case where a device can communicate with an (asymptotically) infinite number of other devices, which reflects the limited communication capability of each device in real world.

The two assumptions are consistent with the extended network model widely used in the analysis of large networks.

### B. Extinction probability

Based on the knowledge of  $\gamma_{h,k}$ , a number of interesting results can be obtained using the theory established in the analysis of epidemiology. Firstly, there is a well-known phase transition phenomenon [33], viz., there exists a threshold in the average node degree below which an infectious disease becomes extinct after some time with probability 1 (almost surely); while above the threshold the disease has a positive probability of spreading out to a large population. Using the analogy in epidemiology, this sub-section provides results on the phase transition phenomenon in the propagation of a packet.

As widely used in the analysis of epidemics, a branching process is used to investigate the propagation of a packet in this section. Specifically, we construct a multi-type branching process [32], [34] to study the number of informed nodes of a typical packet, say packet  $j$ , where the 0<sup>th</sup> generation includes the nodes that receive the packet  $j$  at time 0 (i.e. in the first phase). Furthermore, the number of type- $k$  children generated by a type- $h$  node is given by a random variable  $\hat{Q}_{h,k}$ . Because nodes move independently of one another, it is evident that  $\hat{Q}_{h,k}$  follows a Binomial distribution  $\text{Bin}(N_k, \gamma_{h,k})$ , where  $\gamma_{h,k}$  is given in the previous sub-section.

Denote by  $\mathcal{X}_\alpha^h$  the number of type- $h$  nodes in the  $\alpha^{\text{th}}$  generation.

**Definition 1** (extinction). The branching process for the number of informed nodes for a typical packet becomes *extinct* if  $\sum_{h=1}^H \mathcal{X}_\alpha^h = 0$  for any  $\alpha \in \{1, 2, 3, \dots\}$ .

Following is a well-known result for the threshold phenomenon of the extinction of a multi-type branching process.

**Lemma 1** (Threshold phenomenon). Define the matrix  $\hat{\Xi} \triangleq \{\mathbb{E}[\hat{Q}_{h,k}] = N_k\gamma_{h,k}; h, k = 1, 2, \dots, H\}$ . Let  $R_q$  be the largest eigenvalue of  $\hat{\Xi}$ . Then the branching process will become extinct with probability 1 if and only if  $R_q \leq 1$ .

*Proof.* This result can be readily obtained by applying Theorem 2 in [32, Chapter V].  $\square$

When  $R_q > 1$ , there is a positive probability that the branching process does not become extinct, i.e. the packet can be disseminated to a significant number of nodes. Note that in a heterogeneous network, the type of source nodes that a branching process is rooted at can have a significant impact on the probability that the branching process becomes extinct.

Denote by  $w_h$  the *extinction probability of a type- $h$  source*, which is defined as the probability that a branching process rooted at a type- $h$  node becomes extinct. The following theorem characterises  $w_h$ .

**Theorem 1** (Extinction probability). Consider an asymptotic network with  $L \rightarrow \infty$  while keeping nodes' density unchanged and reducing the node's communication range so that Assumptions 1 and 2 are satisfied. The extinction probabilities  $w_h$  for  $h = 1, 2, \dots, H$  are the solutions to the following system of equations:

$$w_h = \exp\left(\sum_{k=1}^H N_k\gamma_{h,k}(w_k - 1)\right), \text{ for } h = 1, 2, \dots, H. \quad (3)$$

*Proof.* Firstly, note that in the construction of the branching process, we consider that every node that a type- $h$  node meets is a susceptible node. As the packet propagates, more and more nodes receive the packet and hence the probability that a type- $h$  node meets an informed node increases. This may decrease the expected number of newly informed nodes generated by an infectious node. However, there is no need to consider the impact of this in the analysis of extinction probability because when analysing the extinction probability, we are only interested in the case that the fraction of recipients is vanishingly small as  $L \rightarrow \infty$  and  $N \rightarrow \infty$ , i.e. the probability that a type- $h$  node meets an informed node is vanishingly small and negligible.

Recall that according to Assumption 2, when we increase  $L \rightarrow \infty$  while keeping nodes' density unchanged, the network has  $N_k \rightarrow \infty$  while the quantity  $N_k\gamma_{h,k}$  remains a finite constant. Consequently, the distribution of  $\hat{Q}_{h,k}$ , i.e. a Binomial distribution  $\text{Bin}(N_k, \gamma_{h,k})$ , approaches a Poisson distribution with an expected value  $N_k\gamma_{h,k}$  [35]. The difference between  $\hat{Q}_{h,k}$  and its Poisson distribution counterpart (denoted by  $Q_{h,k}$ ) diminishes as  $N_k \rightarrow \infty$  and  $\gamma_{h,k} \rightarrow 0$ , where the convergence rate is given in [35].

Denote by  $G_{h,k}(s)$  the probability generating function of  $Q_{h,k}$ . Because  $Q_{h,k}$  follows a Poisson distribution, it is straightforward that

$$G_{h,k}(s) = \mathbb{E}[s^{Q_{h,k}}] = \exp(N_k\gamma_{h,k}(s - 1)). \quad (4)$$

Furthermore, define the multi-variate probability generating function  $\hat{G}_h(\mathbf{s}) \triangleq \mathbb{E}[s_1^{Q_{h1}} s_2^{Q_{h2}} \dots s_H^{Q_{hH}}]$ , where  $\mathbf{s} \triangleq \{s_1, s_2, \dots, s_H\}$  is a row vector. It can be shown that

$$\begin{aligned} \hat{G}_h(\mathbf{s}) &= \sum_{q_{h1}=0}^{\infty} \sum_{q_{h2}=0}^{\infty} \dots \sum_{q_{hH}=0}^{\infty} \Pr(Q_{h1} = q_{h1}) \Pr(Q_{h2} = q_{h2}) \\ &\quad \dots \Pr(Q_{hH} = q_{hH}) s_1^{q_{h1}} s_2^{q_{h2}} \dots s_H^{q_{hH}} \\ &= G_{h,1}(s_1) G_{h,2}(s_2) \dots G_{h,H}(s_H). \end{aligned} \quad (5)$$

Denote the extinction probabilities by a row vector  $\mathbf{w} \triangleq \{w_1, w_2, \dots, w_H\}$ . Then according to Theorem 2 in [32, Chapter V], if Assumption 1 is satisfied, the extinction probabilities satisfy  $\mathbf{w} = \hat{\mathbf{G}}(\mathbf{w})$ , where  $\hat{\mathbf{G}}(\mathbf{s})$  is a row vector  $(\hat{G}_1(\mathbf{s}), \hat{G}_2(\mathbf{s}), \dots, \hat{G}_H(\mathbf{s}))$ . The conclusion follows that the extinction probabilities  $w_h$  for  $h = \{1, 2, \dots, H\}$  are the solutions

to the following system of equations:

$$\begin{aligned} w_1 &= \prod_{k=1}^H \exp(N_k \gamma_{1,k} (w_k - 1)), \\ &\dots \\ w_H &= \prod_{k=1}^H \exp(N_k \gamma_{H,k} (w_k - 1)), \end{aligned} \quad (6)$$

where  $N_k$  is the number of type- $k$  nodes in the network and  $\gamma_{h,k}$  is given by Eq. 2.  $\square$

Using Theorem 1, we can further obtain the extinction probability for several special cases.

**Corollary 2.** [Extinction probability for homogeneous networks] Consider the special case of a network with only  $H = 1$  type of nodes. Then the extinction probability is  $w_1 = -\frac{\mathcal{W}(-N\gamma_{1,1} \exp(-N\gamma_{1,1}))}{N\gamma_{1,1}}$ , where  $\mathcal{W}(\cdot)$  is the Lambert-W function.

*Proof.* In a network with only  $H = 1$  type of nodes, it can be readily shown that the extinction probability is the solution to  $w_1 = \exp(N\gamma_{1,1}(w_1 - 1))$  using Theorem 1. Solving this equation straightforwardly yields the unique value  $w_1 = -\frac{\mathcal{W}(-N\gamma_{1,1} \exp(-N\gamma_{1,1}))}{N\gamma_{1,1}}$ , where  $\mathcal{W}(\cdot)$  is the Lambert-W function [36].  $\square$

**Corollary 3.** [Extinction probability with multiple source nodes] Suppose that a packet is initially disseminated from  $\beta = \sum_{h=1}^H \beta_h$  source nodes at the beginning of phase 2, where  $\beta_h$  is the number of type- $h$  source nodes. Then the extinction probability for this packet is  $\prod_{h=1}^H w_h^{\beta_h}$ .

*Proof.* Because every node moves and transmits independently of other nodes, each of the type- $h$  child nodes generated by a parent node becomes an ancestor of an independent subprocesses (which restarts with the type- $h$  node) [37]. Because of this so-called "multiplicative" nature [37], the probability that all the branching processes, rooted at  $\beta$  initial nodes, become extinct is  $\prod_{h=1}^H w_h^{\beta_h}$ .  $\square$

Note that when the branching process does *not* become extinct, the packet is disseminated to a significant number of nodes and we say that *the packet spreads out*. The next subsection quantifies the number of recipients of a packet when it spreads out.

### C. Expected fraction of recipients

Denote by  $\hat{z}_h$  the expected fraction of type- $h$  nodes which receive the packet  $j$  in the steady state, where the packet is initially broadcast from a randomly chosen source node and the packet spreads out. In this subsection, we further investigate  $\hat{z}_h$ .

**Theorem 2** (Fraction of recipients when a packet spreads out). In a large network with  $N \rightarrow \infty$ , suppose that the packet  $j$  is disseminated from a randomly chosen source node of arbitrary type. Given that the packet is spread out, the expected fractions of recipients  $\hat{z}_h$  for  $h = 1, 2, \dots, H$  are the solutions to the

following system of equations

$$1 - \hat{z}_h = \exp\left(-\sum_{k=1}^H N_k \gamma_{k,h} \hat{z}_k\right), \text{ for } h = 1, 2, \dots, H. \quad (7)$$

*Proof.* According to the analysis of epidemics [33], [38], conditioned on the epidemic not becoming extinct and under the assumption that the initial number of infected nodes is relatively small compared with the total number of nodes in the network, the probability of a type- $h$  node not being infected is given by the unique positive solution to the following equation

$$\begin{aligned} 1 - \hat{z}_h &= \prod_{k=1}^H \exp(-N_k \gamma_{k,h} \hat{z}_k) \\ &= \exp\left(-\sum_{k=1}^H N_k \gamma_{k,h} \hat{z}_k\right), \text{ for } h = 1, 2, \dots, H, \end{aligned} \quad (8)$$

where  $\gamma_{k,h}$  is the probability that the type- $h$  node is infected by a type- $k$  node during the type- $k$  node's active period, given by Eq. 2, and  $N_k \hat{z}_k$  is the total expected number of type- $k$  nodes that have been infected during the whole epidemic sharing process.  $\square$

Next we consider the case that there is more than one source nodes of the packet  $j$ .

**Corollary 4** (Fraction of recipients with multiple source nodes). Suppose that at the initial time 0, there are  $\sum_{h=1}^H \beta_h = \beta$  nodes that have received the packet  $j$ , where  $\beta_h$  is the number of type- $h$  source nodes. Denote by  $z(\beta_1, \dots, \beta_H)$  the expected fraction of nodes, out of the total  $N$  nodes, which receive the packet in the steady state. In a large network with  $N \rightarrow \infty$ , there holds

$$z(\beta_1, \dots, \beta_H) = \left(\sum_{h=1}^H \frac{N_h \hat{z}_h}{N}\right) \left(1 - \prod_{h=1}^H w_h^{\beta_h}\right). \quad (9)$$

*Proof.* From Corollary 3, the probability that the packet spreads out is  $(1 - \prod_{h=1}^H w_h^{\beta_h})$ . Note that if the packet does not spread out, the fraction of recipients goes to 0 as  $N \rightarrow \infty$ . If the packet spreads out, the expected number of type- $h$  recipients is  $N_h \hat{z}_h$ , where  $\hat{z}_h$  is given by Theorem 2. The conclusion follows.  $\square$

For the special case that there is only  $H = 1$  type of node in a network, a closed form expression of the fraction of informed nodes in the steady state can be obtained.

**Corollary 5** (Fraction of recipients for homogeneous networks). Suppose that there is only  $H = 1$  type of nodes in a network and a packet is sent to  $\beta$  different nodes in the first phase. Then in the steady state, the expected fraction of recipients of this packet is

$$\begin{aligned} z(\beta) &= \left(1 + \frac{\mathcal{W}(-N\gamma_{1,1} \exp(-N\gamma_{1,1}))}{N\gamma_{1,1}}\right) \\ &\quad \left(1 - \left(-\frac{\mathcal{W}(-N\gamma_{1,1} \exp(-N\gamma_{1,1}))}{N\gamma_{1,1}}\right)^\beta\right). \end{aligned} \quad (10)$$

*Proof.* When  $H = 1$ , Eq. 7 becomes  $1 - \hat{z}_1 = \exp(-N\gamma_{1,1} \hat{z}_1)$ , whose solution is  $\hat{z}_1 = 1 + \frac{\mathcal{W}(-N\gamma_{1,1} \exp(-N\gamma_{1,1}))}{N\gamma_{1,1}}$ . Then Eq. 10 can be easily obtained using Corollary 4.  $\square$

It is interesting to note that when  $\beta = 1$ , the expected fraction of recipients of a packet, given that the packet spreads out, is equal to the probability that the packet spreads out. This coincides with the previous results in single-type epidemic [39] that if the active period is a constant value, the expected final fraction of recipients for an epidemic, given that it does not become extinct, is equal to the probability that the epidemic does not become extinct. On the other hand, when  $H > 1$ , these two quantities can be significantly different.

#### D. Duration of the sharing phase

To fully utilize the complementary network to offload data traffic, the duration of the sharing phase, i.e.  $\mathcal{T}_{end}$ , needs to be longer than the duration of the epidemic sharing process. This section proposes two methods to determine  $\mathcal{T}_{end}$ .

The first method is based on the report of the node to the BSs. Each time a node receives a new packet, i.e. a node starts a new infectious period, the node transmits a short notice (the size of the notice is much smaller than the size of data packets) to the BSs so that the BSs can calculate the time when the epidemic sharing stops. For example, if the last notice is received at time  $t_l$  from a type- $h$  node, then the sharing phase stops at time  $\mathcal{T}_{end} = t_l + \tau_h$ .

Separately,  $\mathcal{T}_{end}$  can also be estimated analytically. The existing research in epidemiology [40, Chapter 7.5] has shown using a Markov Chain model that the duration of a single type ( $H = 1$ ) SIR epidemic can be approximately calculated as follows.

Denote by  $\mathcal{T}_{i,j}$  the expected duration of an SIR epidemic given that there are  $i$  susceptible nodes and  $j$  infectious nodes. Then according to [40, Chapter 7.5],  $\mathcal{T}_{i,j}$  satisfies

$$\mathcal{T}_{i,j} = \frac{1}{j/\tau + ij/\lambda} + p_i \mathcal{T}_{i,j-1} + (1 - p_i) \mathcal{T}_{i-1,j+1} \quad (11)$$

where  $p_i = \frac{\lambda}{\lambda + i\tau}$  is the probability that an infectious node recovers and  $1 - p_i$  is the probability that a susceptible node is infected.

With initial condition  $\mathcal{T}_{i,0} = 0$  for any value of  $i$ , the above system of equations can be numerically solved.

#### V. MINIMISING THE TRAFFIC LOAD OF CELLULAR NETWORKS

Based on the above characterisation of the content dissemination process, this section investigates the optimal content dissemination strategy that minimises the total traffic load of cellular networks.

Recall that in the initial phase, the BSs transmit  $\beta$  packets to  $\beta$  different nodes through the cellular network. Denote by  $Y$  the expected number of packets that BSs need to transmit in the complement phase.

**Definition 2** (Cellular traffic load). The *cellular traffic load* is the expected number of packets transmitted by BSs through the cellular network, which consist of the packets transmitted in the first and the third phases, i.e.  $\beta + Y$ , in order to transmit  $M$  messages to  $N$  nodes.

Note that the value of  $\beta$  determines the value of  $Y$ , which is calculated later in this section. Then the problem of minimising the cellular traffic load can be formulated as follows:

$$\begin{aligned} & \text{Minimise} && \beta + Y \\ & && \beta \\ & \text{Subject to} && \beta \in \{1, 2, \dots, N\}. \end{aligned} \quad (12)$$

##### A. Case 1 - Without coding

This sub-section considers the case that no coding technique is employed for comparison with our proposed dissemination strategy employing coding. Recall that without coding, each of the  $\beta$  packets transmitted in the first phase contains one of the  $M$  messages. As described in Section III-C, among the  $\beta$  packets, there are  $\hat{\beta}_m \in \{0, 1, \dots, \beta\}$  copies of the  $m^{\text{th}}$  message for  $m = \{1, 2, \dots, M\}$ . Further, denote by  $\hat{\beta}_{m,h} \in \{0, 1, \dots, \beta_m\}$  the number of type- $h$  nodes that receive the  $m^{\text{th}}$  message in the first phase. Note that  $\sum_{h=1}^H \hat{\beta}_{m,h} = \hat{\beta}_m$  and  $\sum_{m=1}^M \hat{\beta}_m = \beta$ .

Recall that each node needs at least one copy of each message by the end of the third phase. It is straightforward that

$$Y = \sum_{m=1}^M N \left(1 - z(\hat{\beta}_{m,1}, \dots, \hat{\beta}_{m,H})\right), \quad (13)$$

where  $z(\hat{\beta}_{m,1}, \dots, \hat{\beta}_{m,H})$  is the expected fraction of recipients given by Corollary 4.

Then the optimisation problem in Eq. 12 can be readily solved. Due to the complexity in the expressions of  $z(\hat{\beta}_{m,1}, \dots, \hat{\beta}_{m,H})$ , the following lemmas only present the analytical solution to the optimisation problem for  $H = 1$ .

We first investigate the optimal values of  $\hat{\beta}_m$  for  $m = \{1, 2, \dots, M\}$  that minimise the cellular traffic load.

**Lemma 6** (Diversity). *If  $H = 1$  and  $\beta \leq M$ , the strategy that minimises the cellular traffic load is to push  $\beta$  different messages in the initial phase rather than pushing multiple copies of the same message.*

*Proof.* If the  $m^{\text{th}}$  message for  $m \in \{1, 2, \dots, M\}$  has  $\hat{\beta}_m$  source nodes, then according to Corollary 5, the expected fraction of recipients of the  $m^{\text{th}}$  message is  $z(\hat{\beta}_m) = (1 - w_1)(1 - w_1^{\hat{\beta}_m})$ , where  $w_1 = -\frac{\mathcal{W}(-N\gamma_{1,1} \exp(-N\gamma_{1,1}))}{N\gamma_{1,1}}$  is given by Corollary 2.

It is straightforward that the number of packets that the BSs need to transmit in the third phase is

$$Y = \sum_{m=1}^M N(1 - z(\hat{\beta}_m)) = \sum_{m=1}^M N(w_1 + w_1^{\hat{\beta}_m}(1 - w_1)). \quad (14)$$

Given  $N$  and  $w_1$ , which are determined by the properties of a network, in order to minimise  $Y$ , one needs to minimise  $\sum_{m=1}^M w_1^{\hat{\beta}_m}$ , i.e.

$$\begin{aligned} & \text{Minimise} && \sum_{m=1}^M w_1^{\hat{\beta}_m} \\ & && \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_M \\ & \text{Subject to} && \sum_{m=1}^M \hat{\beta}_m = \beta \leq M. \end{aligned} \quad (15)$$

Generally this is a multivariate integer programming problem which can be complicated to solve. However, in this particular case, a simple observation can straightforwardly lead



to the solution. Without loss of generality, we start with an arbitrary set of values of  $\hat{\beta}_m$  for  $m = \{1, 2, \dots, M\}$ . Suppose that there exists a  $m_1$  where  $\hat{\beta}_{m_1} > 1$ , then there must exist a  $m_2$  where  $\hat{\beta}_{m_2} = 0$  because  $\sum_{m=1}^M \hat{\beta}_m = \beta \leq M$ . Now if we decrease  $\hat{\beta}_{m_1}$  by one, the objective  $\sum_{m=1}^M w_1^{\hat{\beta}_m}$  is increased by  $w_1^{\hat{\beta}_{m_1}-1} - w_1^{\hat{\beta}_{m_1}} = w_1^{\hat{\beta}_{m_1}-1}(1 - w_1)$ . At the same time, in order to satisfy the constraint  $\sum_{m=1}^M \hat{\beta}_m = \beta$ , we increase  $\hat{\beta}_{m_2}$  from zero to one. This step causes the objective  $\sum_{m=1}^M w_1^{\hat{\beta}_m}$  to be decreased by  $1 - w_1$ . Because  $0 \leq w_1 \leq 1$  and  $\hat{\beta}_{m_1} > 1$ , there holds  $1 - w_1 \geq w_1^{\hat{\beta}_{m_1}-1}(1 - w_1)$ . Therefore, these two steps decrease the objective  $\sum_{m=1}^M w_1^{\hat{\beta}_m}$ . Following the same procedure, one can easily reach the conclusion that  $\sum_{m=1}^M w_1^{\hat{\beta}_m}$  is minimised when  $\hat{\beta}_m \leq 1$  for every  $m = \{1, 2, \dots, M\}$ .  $\square$

Based on Lemma 6, the following lemma can be obtained.

**Lemma 7** (Diversity maximizing strategy). *Suppose that the BSs push  $\beta$  packets to  $\beta$  different nodes, where  $\beta$  is in the range  $cM < \beta \leq (c+1)M$  for a constant  $c \in \{0, 1, 2, \dots\}$ . Then the optimal strategy that minimises the cellular traffic load is to push  $\hat{\beta}_m$  copies of the  $m^{\text{th}}$  message where  $c \leq \hat{\beta}_m \leq c+1$  for every  $m = \{1, 2, \dots, M\}$  and  $\sum_{m=1}^M \hat{\beta}_m = \beta$ .*

*Proof.* According to Lemma 6, it is straightforward that the first  $M$  packets, among the total  $\beta$  packets pushed in the first phase, should contain  $M$  different messages. Using the same technique as that in the proof of Lemma 6, it can be shown that the optimal strategy is the one illustrated in Fig. 2.

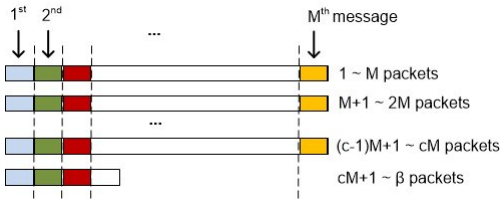


Fig. 2. Illustration of the strategy of allocating  $M$  messages in  $\beta$  packets. The packets are arranged in  $c+1$  rows, where each of the first  $c$  rows contain  $M$  packets and the last row contains  $\beta - cM$  packets. According to Lemma 7, each message has at least  $c$  copies and at most  $c+1$  copies.  $\square$

After investigating the packets pushed in the first phase, we next calculate the number of packets pulled in the third phase, denoted by  $Y$ . Suppose that  $cM < \beta \leq (c+1)M$  for  $c \in \{0, 1, 2, \dots\}$ , then according to Lemma 7 there are  $\beta - cM$  messages where each one has  $c+1$  source nodes, and the number of source nodes for each of the remaining  $M - \beta + cM$  messages is  $c$ .

Then according to Eq. 14, the expected number of packets requiring to be transmitted over cellular networks in the third phase can be readily obtained:

$$Y = (\beta - cM)N(w_1 + w_1^{c+1}(1 - w_1)) + (M - \beta + cM)N(w_1 + w_1^c(1 - w_1)). \quad (16)$$

With the above expression of  $Y$ , the optimisation problem specified in Eq. 12 can be readily solved by Matlab.

The following theorem provides a closed-form solution to the optimal number of initial copies transmitted through cellular networks in the first phase for the special case of a homogeneous network with one message only.

**Theorem 3** (Optimisation). *In a network with only  $H = 1$  type of nodes and  $M = 1$  message, the cellular traffic load is minimised when*

$$\beta = \log_{w_1} \left( \frac{1}{N(w_1 - 1) \ln w_1} \right), \quad (17)$$

where  $w_1 = -\frac{\mathcal{W}(-N\gamma_{1,1} \exp(-N\gamma_{1,1}))}{N\gamma_{1,1}}$  is given by Corollary 2.

*Proof.* According to Lemma 6, when  $H = 1$  and  $M = 1$ , there holds  $\beta + Y = \beta + N(1 - z(\beta))$ , where  $z(\beta)$  is the expected fraction of recipients given by Corollary 4. Then the problem of minimising cellular traffic load can be formulated as follows

$$\begin{aligned} & \underset{\beta}{\text{Minimise}} && \beta + N(1 - (1 - w_1)(1 - w_1^\beta)) \\ & \text{Subject to} && \beta \in \{1, 2, \dots, N\}. \end{aligned} \quad (18)$$

Take the second derivative of the objective function, there yields

$$\begin{aligned} & \frac{\partial^2}{\partial \beta^2} (\beta + N(1 - (1 - w_1)(1 - w_1^\beta))) \\ & = N(1 - w_1)w_1^\beta \ln^2(w_1) \geq 0. \end{aligned} \quad (19)$$

Therefore, the objective function is convex. Further, the first derivative of the objective function is

$$\begin{aligned} & \frac{\partial}{\partial \beta} (\beta + N(1 - (1 - w_1)(1 - w_1^\beta))) \\ & = 1 + N(1 - w_1)w_1^\beta \ln(w_1), \end{aligned} \quad (20)$$

which is equal to zero when  $\beta = \log_{w_1} \left( \frac{1}{N(w_1 - 1) \ln w_1} \right)$ .  $\square$

### B. Case 2 - with coding

Using the erasure coding technique introduced in Section III-C2, BSs push  $\beta$  coded packets to nodes in the initial phase. Lemma 8 gives the optimum strategy to select the nodes to distribute the  $\beta$  coded packets in the initial phase assuming that BSs push  $\beta$  encoded packets to  $\beta$  different nodes.

**Lemma 8** (Sharing maximising strategy). *Label all nodes in the network in the descending order of their values of  $w_h$ , which is given by Theorem 1. If more than one node have the same value of  $w_h$ , their order can be arbitrarily assigned. Suppose that BSs push  $\beta$  encoded packets to  $\beta$  different nodes in the initial phase. Then the optimal strategy that minimises the cellular traffic load is to push  $\beta$  different packets to the first  $\beta$  nodes in the above order.*

*Proof.* Using the same technique as that in Lemma 6, it can be easily shown that the strategy that minimises the cellular traffic load is to push  $\beta$  different coded packets in the initial phase rather than pushing multiple copies of the same coded packet.

Because different encoded packets are shared independently of one another, we next consider a randomly chosen encoded packet, say packet  $j$ . It is obvious that to minimise the cellular

traffic load, one needs to maximise the number of recipients of packet  $j$ . According to Corollary 4, the expected fraction of nodes, out of the total  $N$  nodes, which receive packet  $j$  in the steady state is

$$z(\beta_1, \dots, \beta_H) = \left( \sum_{h=1}^H \frac{N_h \hat{z}_h}{N} \right) \left( 1 - \prod_{h=1}^H w_h^{\beta_h} \right), \quad (21)$$

where  $\beta_1 + \beta_2 + \dots + \beta_H = 1$  as a consequence of the discussion in the previous paragraph. In other words, only one value among  $\beta_1, \beta_2, \dots, \beta_H$  is equal to one and the other values are all equal to 0. It is obvious that to maximise  $z(\beta_1, \dots, \beta_H)$ , one should assign the values of  $\beta_1, \beta_2, \dots, \beta_H$  in a way that minimises  $\prod_{h=1}^H w_h^{\beta_h}$ , i.e. let the only non-zero value  $\beta_k = 1$  for the  $k^{\text{th}}$  type of nodes that have the smallest value of the extinction probability  $w_k$  among all  $w_h$  for  $h \in \{1, 2, \dots, H\}$ .

Since BSs push  $\beta$  encoded packets to  $\beta$  different nodes in the initial phase, when the number of type- $k$  nodes  $N_k$  is less than the total number of packets  $\beta$ , some packets need to be pushed to the nodes that have the second (and if needed, the third, fourth, etc.) smallest value of the extinction probability  $w_i$  among all  $w_h$  for  $h \in \{1, 2, \dots, H\}$ .  $\square$

In Lemma 8, the restriction that BSs push  $\beta$  encoded packets to  $\beta$  different nodes in the initial phase is mainly employed for fairness and to avoid overloading a single device. When these considerations are of minor consequences, i.e. BSs can push multiple encoded packets into the same node, the optimum strategy to distribute the  $\beta$  encoded packets follows readily from the analysis presented in the proof of Lemma 8.

Next we study the optimum value of  $\beta$  for the special case of a homogeneous network with  $H = 1$ . The optimum value of  $\beta$  for the more general case of a heterogeneous network with  $H > 1$  can be determined analogously albeit with greater complexity.

Denote by random variable  $B$  the number of packets received by a randomly chosen node at the end of the sharing phase. Then  $B$  follows a Binomial distribution, i.e. the probability that the node receives  $B = b$  packets is  $\Pr(B = b) = \binom{\beta}{b} (z(1))^b (1 - z(1))^{\beta-b}$ , where  $\binom{\beta}{b} = \frac{\beta!}{b!(\beta-b)!}$ . It follows that in the complement phase, the number of packets that need to be transmitted to a randomly chosen node is  $(M - B)^+$ , where  $(x)^+ = \max\{0, x\}$ .

Finally, the expected number of packets that the BSs need to transmit in the complement phases is

$$Y = NE[(M - B)^+] \quad (22)$$

$$= N \sum_{b=0}^M (M - b) \binom{\beta}{b} (z(1))^b (1 - z(1))^{\beta-b}. \quad (23)$$

Then the optimization problem in Eq. 12 becomes

$$\text{Minimise}_{\beta} \quad \beta + N \sum_{b=0}^M (M - b) \binom{\beta}{b} z_1^b (1 - z_1)^{\beta-b} \quad (24)$$

$$\text{Subject to} \quad \beta \in \{1, 2, \dots, N\}.$$

This is can be easily solved using Matlab, where the results are presented in the next section.

*Remark 1.* As shown in this section,  $\beta$  is an important parameter that can be optimised to reduce the traffic load of cellular networks. The best choice of  $\beta$  that minimises the traffic load of cellular networks depends on a number of parameters such as the density of nodes, the number of types of nodes and the nodal moving speed. The number of messages  $M$  is also an important parameter affecting the selection of  $\beta$ . As shown in this section, more specifically in Eq. 24, the value of  $\beta$  that minimises the traffic load of cellular networks depends on the value of  $M$ . Further, Fig. 2 also explains the strategy to generate  $\beta$  packets from  $M$  messages for different values of  $M$  and without employing coding.

## VI. SIMULATION AND DISCUSSION

This section reports on simulations to verify the accuracy of the analysis presented in the previous sections. The simulations are conducted using a mobile network simulator written in C++. Specifically,  $N = 960$  nodes are uniformly deployed on a torus  $(0, 8000)^2$ . Consequently the nodes' density equals to 15 nodes/km<sup>2</sup>, which is the density of cabs in New York metropolitan area [41]. After initial deployment of the nodes, they start to move according to the random direction mobility model introduced in Section III. Unless otherwise specified, the node's speed is  $V = 10$  (typical vehicle moving speed [42]). The radio range  $r_0 = 20\text{m}$  or  $250\text{m}$  (typical radio ranges using Wi-Fi Tethering [10] or DSRC [8]). Every point shown in the simulation result is the average value from 500 simulations.

### A. Exponential inter-meeting model

We first calculate the parameters characterising the exponential distribution of inter-meeting time.

Consider two types of nodes moving according to the random direction model with speeds  $V_1 = 10$  and  $V_2 = 0$ , i.e. mobile and static nodes. The relative speeds between nodes (where  $\tilde{V}_{i,j}$  denotes the relative speed between a type  $i$  node and a type  $j$  node) can be easily calculated (c.f. Lemma 9 in the Appendix):  $\tilde{V}_{1,2} = 10$ ,  $\tilde{V}_{2,2} = 0$  and

$$\tilde{V}_{1,1} = \int_0^{2\pi} \sqrt{10^2 + 10^2 - 200 \cos \theta} \frac{1}{2\pi} d\theta = \frac{40}{\pi}. \quad (25)$$

Consequently, the inter-meeting time between nodes follows an exponential distribution with mean (c.f. Lemma 10 in the Appendix)  $\lambda_{1,1} = \frac{2r_0\tilde{V}}{L^2} = \frac{80r_0}{\pi L^2}$ ,  $\lambda_{1,2} = \frac{2r_0\tilde{V}}{L^2} = \frac{20r_0}{L^2}$  and  $\lambda_{2,2} = 0$  respectively.

Fig. 3 verifies the accuracy of the results of the exponential inter-meeting time distribution given in Lemma 10.

### B. Performance of the sharing phase

Fig. 4 shows the probability that a packet spreads out and the expected fraction of recipients of a single packet in a network with  $H = 2$  types of nodes. The network setting is the same as that used in the previous sub-section. Specifically, half of the nodes (type-1) are mobile nodes and the other half of the nodes (type-2) are static nodes. In Fig. 4(a), the analytical result of the probability that a packet spreads out

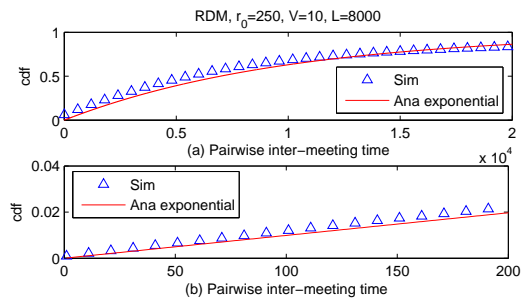


Fig. 3. Verifying exponential inter-meeting time distribution. *Sim* represents the simulation result of the inter-meeting time between two nodes moving under the random direction model. *Ana exponential* represents the analytical result given in Lemma 10. Note that subplot (b) shows the first 200 seconds of subplot (a).

when the source node is of type-1 (resp. type-2) is given by  $w_1$  (resp.  $w_2$ ) from Theorem 1. In Fig. 4(b), the analytical result of the expected fraction of recipients of a packet when the source node is of type-1 (resp. type-2) is given by  $z(1,0)$  (resp.  $z(0,1)$ ) from Corollary 4. It is interesting to note that the probability that a packet spreads out and the expected fraction of recipients can be significantly affected by the type of source node. Specifically, a mobile source node can spread the packet to more recipients than a static source node.

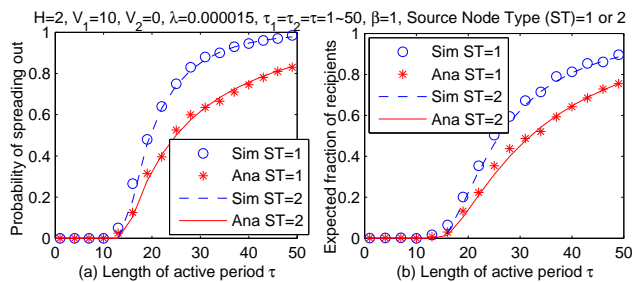


Fig. 4. Simulation and analytical results of (a) The probability that a packet spreads out and (b) the expected fraction of recipients of a packet.

Fig. 5 shows the results of another interesting case where the complementary network consists of a set of fixed-location base stations (e.g. WiFi APs). In the second phase, the message is disseminated from these WiFi APs to mobile users (i.e. type-1 nodes). Specifically, we consider that a small set of WiFi APs are randomly and uniformly deployed in a given area, which is a widely-used setting for AP deployment [3]. Because the WiFi APs are usually connected to the Internet via wired connections, we set  $\gamma_{2,2} = 1$ . Other parameters are the same as those in the previous sub-section. A message is transmitted to all the WiFi APs at time 0. Then the WiFi APs keep transmitting the message for a given time period  $\tau_2$ . In Fig. 5, we let the active period of type-2 nodes be  $\tau_2 = 500, 1500$  respectively while varying the active period of type-1 nodes. Note that  $\tau_1 = 0$  corresponds to the traditional case [4], [9] where nodes do not cooperatively share received packets and they solely rely on WiFi APs to offload data traffic from cellular networks. It can be seen in Fig. 5 that a longer active period of mobile nodes  $\tau_1$  leads to a larger expected fraction of recipients. It is obvious that packet sharing using ad hoc connections between mobile nodes can significantly

increase the number of recipients of a packet, hence reducing the number of transmissions required by BSs.

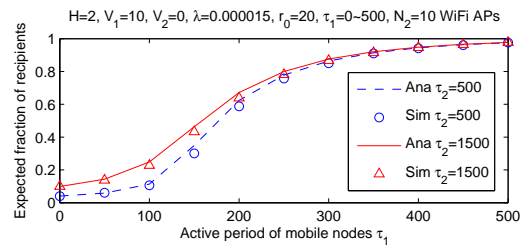


Fig. 5. The expected fraction of recipients of a packet in a network with 10 fixed WiFi APs and some mobile nodes.

Fig. 6 shows the time duration of the epidemic sharing process, i.e. the time required for the epidemic sharing process to reach its steady state, using the same network setting as that in the previous figure. It is obvious that when the active period of mobile nodes is  $\tau_1 = 0$ , the time duration of the sharing phase is equal to the active period of the WiFi APs. Then the time duration of the sharing phase increases as  $\tau_1$  increases. This is because a longer active period of mobile nodes  $\tau_1$  leads to more infected mobile nodes hence a longer time duration for the sharing phase. Above a certain value of  $\tau_1$ , e.g. 300 in this case, an increase in  $\tau_1$  has a limited impact on the time duration of the sharing phase because there is limited increase in the fraction of infected nodes above  $\tau_1 = 300$  as can be seen in Fig. 5 and most nodes are already infected in the first 1000 seconds of the sharing phase. This motivates further work to study the benefit of terminating the epidemic sharing phase before it reaches the steady state.

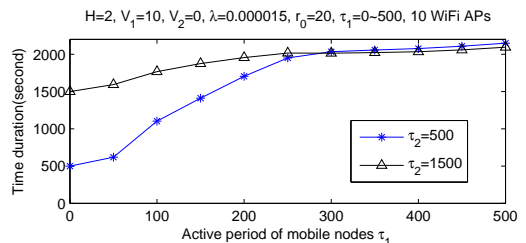


Fig. 6. Simulation results of the time duration of the epidemic sharing phase.

Fig. 7 shows the probability that a packet spreads out and the expected fraction of recipients of a single packet in a network with  $H = 3$  types of nodes, where the proportion of nodes moving at speed  $20m/s$ ,  $10m/s$  and  $0m/s$  is 25%, 25% and 50% respectively. Besides the fact that the analytical results well match the simulation results, it is interesting to note that the choice of the source node type leads to a significant difference in the results. More specifically, it is evident from the curves (and to be expected) that choosing a source node moving at faster speed leads to a higher probability of spreading out and a higher expected fraction of recipients.

Fig. 8 shows the simulation results in a network where nodes move according to the random waypoint mobility model. More specifically, under a random waypoint model [20], a node randomly chooses a destination point in the area and moves at a constant speed along a straight line to this point. After

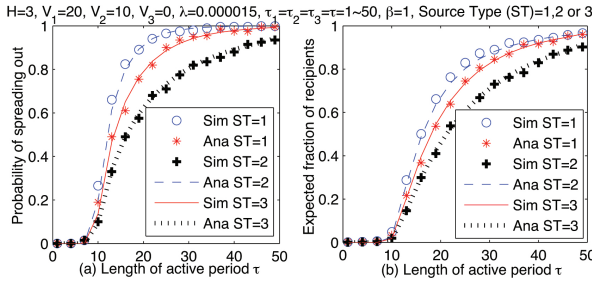


Fig. 7. Simulation and analytical results of (a) The probability that a packet spreads out and (b) the expected fraction of recipients of a packet in a network with  $H = 3$  types of nodes.

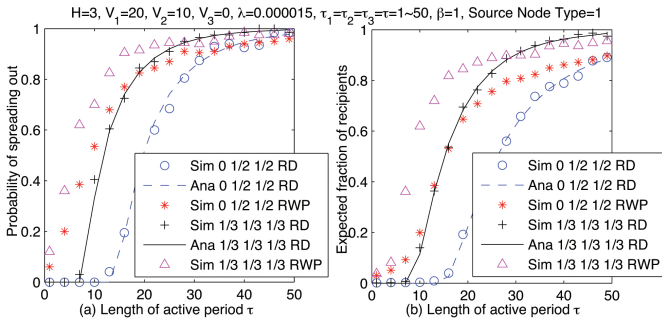


Fig. 8. Simulation and analytical results of (a) The probability that a packet spreads out and (b) the expected fraction of recipients of a packet in a network with  $H = 3$  types of nodes, where nodes are moving according to random direction model (RD) and random waypoint model (RWP).

reaching the point, the node repeats the above process. Fig. 8 shows the probability that a packet spreads out and the expected fraction of recipients of a single packet in a network with  $H = 3$  types of nodes, where the proportion of nodes moving at speed  $20m/s$ ,  $10m/s$  and  $0m/s$  is shown in the legend. It can be seen that the probability of spreading out and the expected fraction of recipients for the networks under the random waypoint model are higher than those under the random direction model. This is because a special property of the random waypoint model - it does not preserve uniform distribution of nodes [18]. More specifically, if nodes are uniformly distributed in an area at time 0 and move according to the random waypoint model, then after some time the density of nodes near the centre of the area will become higher than that near the boundary of the area, i.e. the spatial distribution of nodes is no longer uniform. This causes our analysis, which is based on the assumption of uniform distribution of nodes, to deviate from the simulation results, though the trend is similar. A more detailed study reveals that due to a higher density of nodes near the centre of the area, the packet spreads out more easily, as can be seen in Fig. 8(a), compared with the network with the random direction model. On the other hand, the nodes near the boundary of the area receive the packet with more difficulty, due to the lower density of nodes near the edge of the area, compared with the network with random direction model; this leads to a lower expected fraction of recipients when  $\tau > 30$  in Fig. 8(b).

### C. Minimising cellular traffic load

Based on the results of the packet sharing in the second phase, this sub-section evaluates the cellular traffic loads.

Fig. 9 shows the results of a basic network with  $M = 1$  message,  $H = 1$  type of nodes where no coding technique is employed. Specifically, Fig. 9(a) shows the expected fraction of recipients of single packet at the end of the sharing phase. The analytical result is calculated using Corollary 5. It can be seen in Fig. 9(a) that an increase in either the active period  $\tau$  or the number of source nodes  $\beta$  leads to an increase in the expected fraction of recipients of the packet. On the other hand, a different trend can be observed in Fig. 9(b), which shows the expected cellular traffic load  $\beta + Y$ , where the analytical result of  $Y$  is calculated using Eq. 16. It can be seen that initially the cellular traffic first decreases rapidly when  $\beta$  increases, due to a rapid increase in the expected fraction of recipients in the sharing phase as shown in Fig. 9(a). Then after  $\beta$  reaching a certain value, the cellular traffic load increases gradually as  $\beta$  further increases. This is because the expected fraction of recipients has limited increase when  $\beta$  increases further; on the other hand the increase in  $\beta$  causes more cellular data traffic. It is interesting to note that sending out more packets in the initial phase is not always beneficial.

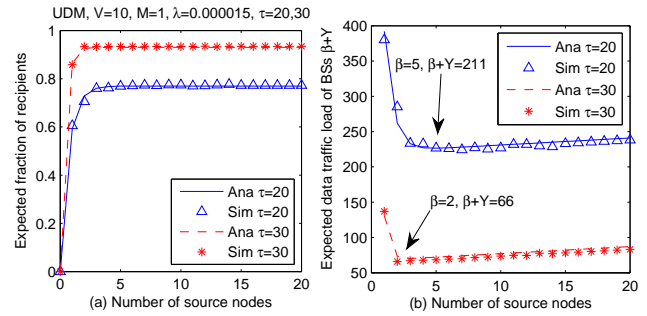


Fig. 9. Results for a network consists of  $H = 1$  type of nodes without employing coding techniques. Specifically, subplot (a) shows the expected fraction of recipients of a single packet and subplot (b) shows the expected cellular data traffic load  $\beta + Y$ .

Fig. 10 shows the expected cellular data traffic load  $\beta + Y$  with different values of  $\beta$ , viz. the number of packets sent in the first phase, using different content dissemination strategies. To study the impact of coding on the cellular traffic load, we consider two networks with the same setting except that one network employs the erasure coding technique (c.f. Section III-C) but the other network does not. Several interesting trends can be observed in Fig. 10. Firstly, when  $\beta$  is small, the two networks have a similar and relatively high cellular traffic load. This is because only a limited number of nodes receive the packets through the complementary network, hence most packets are directly transmitted to the users via cellular networks. As  $\beta$  increases, the cellular traffic load first decreases rapidly then increases gradually, due to the same reason as that in Fig. 9(b).

It is interesting to note that above a certain value of  $\beta$ , e.g.  $\beta = 17$  in Fig. 9(a), the traffic load of cellular networks employing coding is significantly smaller than that of networks without coding, due to the following reason. Recall that when

$\beta > M$ , each node only needs  $M$  different coded packets to decode all  $M$  messages when coding is employed. On the other hand, in a network without coding, there is a non-zero probability that two packets received at a node contain the same message. Therefore a node may need more than  $M$  packets in order to receive all  $M$  messages.

Note that without employing the cooperative content dissemination strategy, the BSs may need to transmit one packet to every one of the 960 nodes via cellular networks, which results in a total of 960 transmissions. Compared with the values of  $\beta + Y$  in Fig. 9 and Fig. 10, e.g. 66 or 18, it is evident that the cooperative content dissemination strategy can significantly reduce the traffic load of cellular networks.

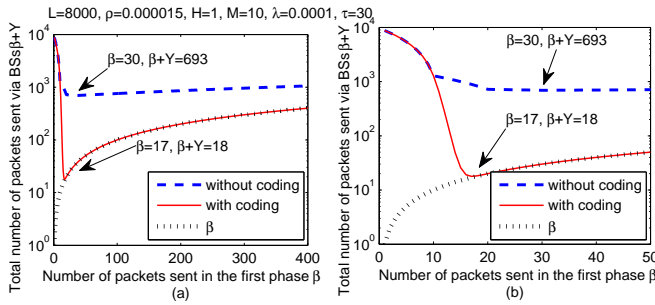


Fig. 10. Comparison between the expected cellular data traffic loads  $\beta + Y$  in networks with and without coding. Note that subplot (b) shows the range  $\beta \in [1, 50]$  of subplot (a).

## VII. CONCLUSION AND FUTURE WORK

This paper proposed a cooperative content dissemination strategy for heterogeneous networks consisting of different types of devices. The content dissemination strategy can effectively offload a significant amount of data traffic from cellular networks to complementary networks such as WiFi and device-to-device networks using ad hoc connections that emerge when devices move and meet one another. A detailed analysis for the content dissemination process was presented, based on which the optimal design of the content dissemination strategy was discussed. The cooperative content dissemination strategy significantly reduces the amount of data traffic from cellular networks while guaranteeing the successful dissemination of the content.

In our paper, we consider that the duration of the sharing phase  $\mathcal{T}_{end}$  is sufficiently long such that the epidemic sharing process is able to reach its steady state. In the future, one may consider the case where only a short time period is allowed for the sharing phase, causing the epidemic sharing process to terminate before it reaches the steady state. In this case, a non-trivial analysis is required to calculate the number of recipients of a packet at an arbitrary time instant. It is also an interesting extension of our work to consider different probability distributions for the inter-meeting time of nodes, where the distribution can be affected by the nodes' mobility and network area.

Our theoretical analysis can only be applied to mobility models for which the resulting spatial distribution of nodes is stationary and ergodic. There is a class of other models, for

example the well known random waypoint mobility model, under which the spatial distribution of nodes is no longer stationary and becomes time-varying. It will be interesting and important to study the performance of our proposed content dissemination strategy under such more general mobility models. Furthermore, our analysis assumes that the spatial distribution of nodes is uniform. Under many scenarios encountered in real applications, the spatial distribution of nodes is more likely to be clustered than uniform. It will also be of interest to study the performance of the proposed dissemination strategy under more general node distributions.

## APPENDIX

In this Appendix, we show that when nodes move according to the well-known random direction mobility model [18], their inter-meeting time distribution is an exponential distribution.

Firstly, we provide the following lemma, which calculates the relative speed between two nodes.

**Lemma 9.** Consider a network where nodes move according to the random direction model, but at different speeds. The expected relative speed between two randomly chosen nodes with known speeds  $V_i$  and  $V_j$  is  $\tilde{V}_{i,j} = \int_0^{2\pi} \sqrt{V_i^2 + V_j^2 - 2V_i V_j \cos \theta} \frac{1}{2\pi} d\theta$ .

*Proof.* Denote by  $\Theta$  the angle measured counterclockwise from the movement direction of a randomly chosen node  $i$  to the movement direction of a randomly chosen node  $j$ . Recall that the direction of a node is randomly and uniformly chosen in  $[0, 2\pi)$ , independent of the directions of other nodes. It follows that  $\Theta$  is uniformly distributed in  $[0, 2\pi)$ .

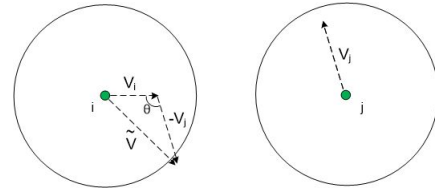


Fig. 11. Illustration of the relative speed between two nodes.

Conditioned on  $\Theta = \theta$ , the relative speed between the two nodes, whose speeds are  $V_i$  and  $V_j$  respectively, is  $\tilde{V}_\theta = \sqrt{V_i^2 + V_j^2 - 2V_i V_j \cos \theta}$ .

Then the expected relative speed can be calculated as

$$\tilde{V}_{i,j} = \int_0^{2\pi} \tilde{V}_\theta \frac{1}{2\pi} d\theta \quad (26)$$

□

Suppose that a node  $h$  is infected at time instant  $t_1$ . Define the first-meeting time interval  $T_{h,k}$  as the time interval from time  $t_1$  to the time when node  $h$  first meets another node  $k$ . The following lemma uses the techniques described in [43, Chapter 4] to estimate the parameters for the distribution of  $T_{h,k}$ .

**Lemma 10.** Denote by  $\tilde{V}_{h,k}$  the relative speed between two nodes  $h$  and  $k$  respectively, both moving in a torus  $(0, L]^2$ . The first-meeting time interval  $T_{h,k}$  approximately follows an

exponential distribution with mean  $\lambda_{h,k} \approx 2r_0\tilde{V}_{h,k}\frac{1}{L^2}$  where  $r_0$  is the radio range of nodes.

*Proof.* Consider a new network where node  $k$  does not move and node  $h$  moves at speed  $\tilde{V}_{h,k}$ . It is straightforward that the distribution of  $T_{h,k}$  in the new network is the same as that in the original network.

Now we consider the new network. Consider a randomly chosen time instant  $t_1$  and assume that node  $h$  is located at  $(x, y)$  at  $t_1$ . The size of the area swept by the radio range of node  $h$  during a differential time period  $\Delta t$  starting from  $t_1$  is  $\Delta A_{x,y} = 2r_0\tilde{V}_{h,k}\Delta t$  [43, Chapter 4].

Recall that at any time instant, the nodes are uniformly distributed in the area. Hence the pdf of node  $k$ 's location is  $\mathcal{P}_k(x, y) = \frac{1}{L^2}$ . Then, the probability that node  $k$  is in the differential area  $\Delta A_{x,y}$  is

$$p_{x,y} = \int \int_{\Delta A_{x,y}} \mathcal{P}_k(x, y) dx dy = \Delta A_{x,y} \frac{1}{L^2}. \quad (27)$$

Similarly, the pdf of node  $h$ 's location at time instant  $t_1$  is  $\mathcal{P}_h(x, y) = \frac{1}{L^2}$ . Then the probability that node  $h$  meets node  $k$  in an incremental time  $\Delta t$  is

$$\begin{aligned} p(\Delta t) &= \int_0^L \int_0^L p_{x,y} \mathcal{P}_h(x, y) dx dy \\ &= \int_0^L \int_0^L 2r_0\tilde{V}_{h,k}\Delta t \frac{1}{L^2} \frac{1}{L^2} dx dy \\ &= 2r_0\tilde{V}_{h,k} \frac{1}{L^2} \Delta t. \end{aligned} \quad (28)$$

Then approximately consider that the event that the two nodes  $h$  and  $k$  meet in one interval of length  $\Delta t$  and the event that the two nodes meet in another non-overlapping interval of length  $\Delta t$  are independent. The accuracy of the approximation is evaluated by simulation in Section VI. It follows that the probability that two nodes do not meet each other within time interval  $t$  is

$$\begin{aligned} \Pr(T_{h,k} > t) &\approx \lim_{\Delta t \rightarrow 0} \prod_{\Delta t \in (0,t)} (1 - p(\Delta t)) \\ &= \lim_{\Delta t \rightarrow 0} \exp\left(-\sum_{\Delta t \in (0,t)} p(\Delta t)\right) \\ &= \exp\left(-2r_0\tilde{V}_{h,k} \frac{1}{L^2} t\right). \end{aligned} \quad (29)$$

It is evident that  $T_{h,k}$  approximately follows an exponential distribution.  $\square$

Note that due to the memoryless property of the exponential distribution, it is straightforward that the inter-meeting between two nodes also follows an exponential distribution.

#### REFERENCES

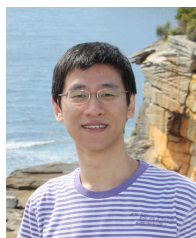
- [1] Y. Li, M. Qian, D. Jin, P. Hui, Z. Wang, and S. Chen, "Multiple mobile data offloading through disruption tolerant networks," *IEEE Transactions on Mobile Computing*, vol. 99, no. PrePrints, pp. 1–1, 2013.
- [2] X. Li, C. Qiao, Y. Hou, and Y. Zhao, "On-road ads delivery scheduling and bandwidth allocation in vehicular cps," in *Proceedings IEEE INFOCOM*, 2013, pp. 2571–2579.
- [3] J. G. Andrews, "Seven ways that hetnets are a cellular paradigm shift," *IEEE Communications Magazine*, vol. 51, no. 3, pp. 136–144, 2013.
- [4] K. Lee, J. Lee, Y. Yi, I. Rhee, and S. Chong, "Mobile data offloading: How much can wifi deliver?" *IEEE/ACM Transactions on Networking*, vol. 21, no. 2, pp. 536–550, 2013.
- [5] M. Taghizadeh, K. Micinski, C. Ofria, E. Torng, and S. Biswas, "Distributed cooperative caching in social wireless networks," *IEEE Transactions on Mobile Computing*, vol. 12, no. 6, pp. 1037–1053, 2013.
- [6] Z. Da and Y. Chai Kiat, "Enabling efficient wifi-based vehicular content distribution," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 3, pp. 479–492, 2013.
- [7] A. Aijaz, H. Aghvami, and M. Amani, "A survey on mobile data offloading: technical and business perspectives," *IEEE Wireless Communications*, vol. 20, no. 2, pp. 104–112, 2013.
- [8] J. Rezgui and S. Cherkaoui, "About deterministic and non-deterministic vehicular communications over dsrc/802.11p," *Wireless Communications and Mobile Computing*, pp. n/a–n/a, 2012.
- [9] E. Bulut and B. K. Szymanski, "Wifi access point deployment for efficient mobile data offloading," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 17, no. 1, pp. 71–78, 2013, 2502948.
- [10] G. Kong and J. K. MBBS, "Using mobile tethering for sharing data across devices: application in rural eye screening," *Journal of Mobile Technology in Medicine (jMTM)*, vol. 1, no. 3, pp. 40–45, 2012.
- [11] H. Bo, H. Pan, V. S. A. Kumar, M. V. Marathe, S. Jianhua, and A. Srinivasan, "Mobile data offloading through opportunistic communications and social participation," *IEEE Transactions on Mobile Computing*, vol. 11, no. 5, pp. 821–834, 2012.
- [12] D. Niyato and E. Hossain, "A unified framework for optimal wireless access for data streaming over vehicle-to-roadside communications," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 6, pp. 3025–3035, 2010.
- [13] S. Panichpapiboon and W. Pattara-Atikom, "A review of information dissemination protocols for vehicular ad hoc networks," *IEEE Communications Surveys and Tutorials*, vol. 14, no. 3, pp. 784–798, 2012.
- [14] C.-H. Lee and D. Y. Eun, "Heterogeneity in contact dynamics: helpful or harmful to forwarding algorithms in dtms?" in *Proceedings IEEE WiOPT*, 2009, pp. 72–81.
- [15] C.-H. Lee and D. Y. Eun, "On the forwarding performance under heterogeneous contact dynamics in mobile opportunistic networks," *IEEE Transactions on Mobile Computing*, vol. 12, no. 6, pp. 1107–1119, 2013.
- [16] L. Yong, W. Zhaocheng, S. Li, J. Depeng, and C. Sheng, "An optimal relaying scheme for delay-tolerant networks with heterogeneous mobile nodes," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 5, pp. 2239–2252, 2013.
- [17] M. Franceschetti and R. Meester, *Random networks for communication: From Statistical Physics to Information Systems*. Cambridge University Press, 2007.
- [18] T. Camp, J. Boleng, and V. Davies, "A survey of mobility models for ad hoc network research," *Wireless Communications and Mobile Computing*, vol. 2, no. 5, pp. 483 – 502, 2002.
- [19] P. Nain, D. Towsley, B. Liu, and Z. Liu, "Properties of random direction models," in *Proceedings IEEE INFOCOM*, vol. 3, 2005, pp. 1897–1907.
- [20] C. Bettstetter, H. Hartenstein, and X. Prez-Costa, "Stochastic properties of the random waypoint mobility model. *wirel. netw.* vol. 10, no. 5, pp. 555–567, 2004." *Wireless Networks*, vol. 10, no. 5, pp. 555–567, 2004.
- [21] T. V. Nguyen and F. Baccelli, "On the spatial modeling of wireless networks by random packing models," in *IEEE INFOCOM*, 2012, pp. 28 – 36.
- [22] G. Mao and B. Anderson, "Capacity of large wireless networks with generally distributed nodes," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1678 – 1691, 2014.
- [23] N. Deng, W. Zhou, and M. Haenggi, "The ginibre point process as a model for wireless networks with repulsion," *IEEE Transactions on Wireless Communications*, vol. 14, no. 1, pp. 107–121, 2015.
- [24] A. Guo and M. Haenggi, "Spatial stochastic models and metrics for the structure of base stations in cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 11, pp. 5800–5812, 2013.
- [25] H. Cai and D. Y. Eun, "Crossing over the bounded domain: From exponential to power-law intermeeting time in mobile ad hoc networks," *IEEE/ACM Transactions on Networking*, vol. 17, no. 5, pp. 1578–1591, 2009.
- [26] M. Sathiamoorthy, A. Dimakis, B. Krishnamachari, and F. Bai, "Distributed storage codes reduce latency in vehicular networks," *IEEE Transactions on Mobile Computing*, vol. 3, no. 9, pp. 2016 – 2027, 2013.
- [27] Y. Li, D. Jin, Z. Wang, L. Zeng, and S. Chen, "Coding or not: Optimal mobile data offloading in opportunistic vehicular networks,"

*IEEE Transactions on Intelligent Transportation Systems*, vol. PP, no. 99, pp. 1–16, 2013.

- [28] Y. Li, P. Hui, D. Jin, L. Su, and L. Zeng, "Evaluating the impact of social selfishness on the epidemic routing in delay tolerant networks," *IEEE Communication Letters*, vol. 14, no. 11, pp. 1026 – 1028, 2010.
- [29] F. D. Pellegrini, D. Miorandi, I. Carreras, and I. Chlamtac, "A graph-based model for disconnected ad hoc networks," in *Proceedings IEEE INFOCOM*, 2007, pp. 373–381.
- [30] X. Zhang, G. Neglia, J. Kurose, and D. Towsley, "Performance modeling of epidemic routing," *Computer Networks: The International Journal of Computer and Telecommunications Networking*, vol. 51, no. 10, pp. 2867–2891, 2007.
- [31] G. Resta and P. Santi, "A framework for routing performance analysis in delay tolerant networks with application to noncooperative networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 1, pp. 2–10, 2012.
- [32] K. B. Athreya and P. E. Ney, *Branching Processes*. Berlin: Springer-Verlag, 1972.
- [33] T. Britton, "Stochastic epidemic models: A survey," *Mathematical Biosciences*, vol. 225, no. 1, pp. 24–35, 2010.
- [34] T. E. Harris, *The Theory of Branching Processes*. Dover Publications, 2002.
- [35] D. Pollard, *Poisson Approximation, Chapter 8, Statistics 241/541: Notes*, 1997.
- [36] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, "On the Lambert W Function," *Advances in Computational Mathematics*, vol. 5, pp. 329–359, 1996.
- [37] K. Mizumoto, K. Ejima, T. Yamamoto, and H. Nishiura, "Vaccination and clinical severity: Is the effectiveness of contact tracing and case isolation hampered by past vaccination?" *International Journal of Environmental Research and Public Health*, vol. 10, no. 3, pp. 816–829, 2013.
- [38] F. Ball and D. Clancy, "The final size and severity of a generalised stochastic multitype epidemic model," *Advances in Applied Probability*, vol. 25, no. 4, pp. 721–736, 1993.
- [39] E. Kenah and J. M. Robins, "Second look at the spread of epidemics on networks," *Physical Review E*, vol. 76, no. 3, pp. 1–12, 2007.
- [40] L. J. S. Allen, *An Introduction to Stochastic Processes with Applications to Biology*, second edition ed. Pearson Education, 2003.
- [41] MKThink, "Unsustainable city: density, transportation, and san franciscos sustainability," 2005. [Online]. Available: <http://www.mkthink.com/archives/2470>
- [42] M. Rudack, M. Meincke, and M. Lott, "On the dynamics of ad hoc networks for inter vehicle communications (IVC)," in *Proceedings ICWN*, 2002.
- [43] R. Groenevelt, "Stochastic models for ad hoc networks," *PhD thesis*, 2005.



**Guoqiang Mao** (S'98-M'02-SM'07) received PhD in telecommunications engineering in 2002 from Edith Cowan University. Between 2002 and 2013, he was an Associate Professor at the School of Electrical and Information Engineering, the University of Sydney. He currently holds the position of Professor of Wireless Networking, Director of Centre for Real-time Information Networks at the University of Technology, Sydney. He has published more than 100 papers in international conferences and journals, which have been cited more than 3000 times. His research interest includes intelligent transport systems, applied graph theory and its applications in networking, wireless multi-hop networks, wireless localisation techniques and network performance analysis. He is an Editor of *IEEE Transactions on Vehicular Technology* and *IEEE Transactions on Wireless Communications*, and a co-chair of *IEEE Intelligent Transport Systems Society Technical Committee on Communication Networks*.



**Zijie Zhang** (S'09-M'12) received his BEng degree in Electronic and Communications Engineering from the University of Hong Kong in 2007. In 2012, he received PhD degree in Engineering from the University of Sydney. Then he worked as a research engineer in National ICT Australia (NICTA). His research interests include graph theoretical analysis and stochastic characterization of wireless communication networks and intelligent transportation systems.



**Brian D. O Anderson** (M'66-SM'74-F'75-LF'07) was born in Sydney, Australia, and educated at Sydney University in mathematics and electrical engineering, with PhD in electrical engineering from Stanford University in 1966. He is a Distinguished Professor at the Australian National University and Distinguished Researcher in National ICT Australia. His awards include the IEEE Control Systems Award of 1997, the 2001 IEEE James H Mulligan, Jr Education Medal, and the Bode Prize of the IEEE Control System Society in 1992, as well as several IEEE and other best paper prizes. He is a Fellow of the Australian Academy of Science, the Australian Academy of Technological Sciences and Engineering, the Royal Society, and a foreign member of the US National Academy of Engineering. He holds honorary doctorates from a number of universities, including Université Catholique de Louvain, Belgium, and ETH, Zürich. He is a past president of the International Federation of Automatic Control and the Australian Academy of Science. His current research interests are in distributed control, sensor networks and econometric modelling.